# A variational method for the Rasch model

Frank Rijmen and Jiří Vomlel

Catholic University Leuven, Belgium
Academy of Sciences of the Czech Republic

Salzburg, December, 3, 2005

# Variables and parameters

## Model variables

| | |
|---|---|
| $Y_{n,i}$ | binary response variable - its values indicates whether the answer of person $n$ to question $i$ was correct |
| $n = 1, \ldots, N$ | person index |
| $i = 1, \ldots, I$ | question index |

## Model parameters

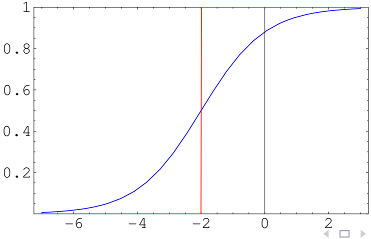
| | |
|---|---|
| $\delta_i$ | difficulty of question $i$ - fixed effects |
| $\beta_n$ | ability (knowledge level) of person $n$ - a random effect |

# Models for the response variable $Y$

$$Y_{n,i} = \begin{cases} 1 & \text{if } \beta_n \geq \delta_i \\ 0 & \text{otherwise.} \end{cases}$$

$$P(Y_{n,i} = 1) = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}$$

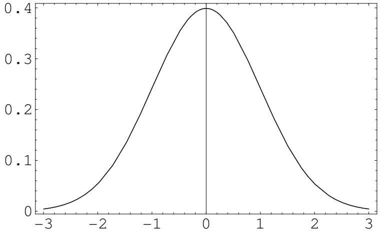
$P(Y_{n,i} = 1 \mid \beta_n)$ for $\delta_i = -2$



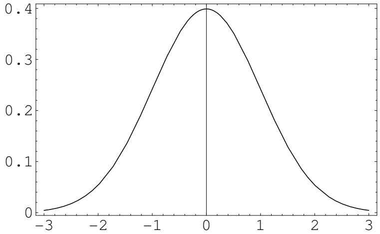
# Probability distribution for random effect $\beta_n$

$$P(\beta_n) \quad = \quad \mathcal{N}(0, \sigma^2)$$

a normal (Gaussian) distribution
with the mean equal zero, and variance $\sigma^2$.

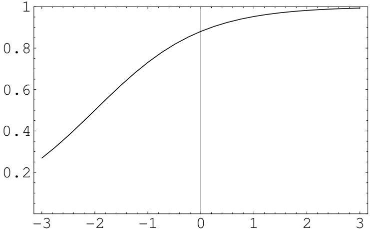


prior $\mathcal{N}_\beta(0,1)$

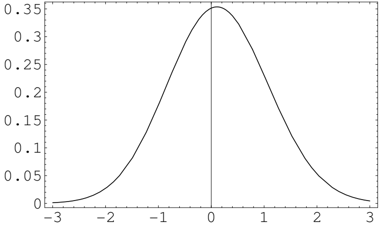



$\mathcal{N}_\beta(0,1)$

$P(Y = 1 \mid \beta, \delta_1 = -2)$



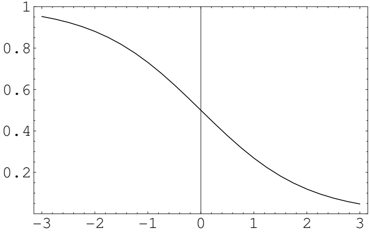

posterior



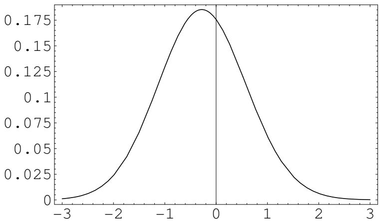

$$\mathcal{N}_\beta(0,1) \cdot P(Y = 1 \mid \beta, \delta_1 = -2)$$

# Computations with the Rasch model

$P(Y = 0 \mid \beta, \delta_2 = 0)$



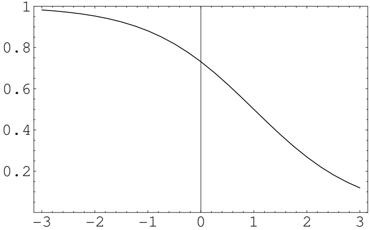

posterior



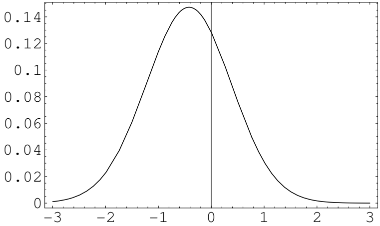

$$\mathcal{N}_\beta(0,1) \cdot P(Y = 1 \mid \beta, \delta_1 = -2) \cdot P(Y = 0 \mid \beta, \delta_2 = 0)$$

# Computations with the Rasch model

$P(Y = 0 \mid \beta, \delta_3 = +1)$



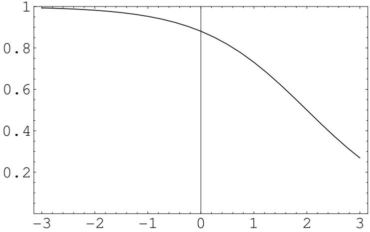

posterior



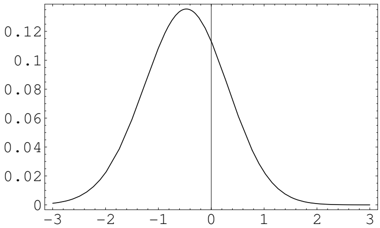

$$\mathcal{N}_\beta(0,1) \cdot P(Y = 1 \mid \beta, \delta_1 = -2) \cdot P(Y = 0 \mid \beta, \delta_2 = 0)$$
$$\cdot P(Y = 0 \mid \beta, \delta_3 = +1)$$

# Computations with the Rasch model

$P(Y = 0 \mid \beta, \delta_4 = +2)$




posterior




$$\mathcal{N}_\beta(0,1) \cdot P(Y = 1 \mid \beta, \delta_1 = -2) \cdot P(Y = 0 \mid \beta, \delta_2 = 0)$$
$$\cdot P(Y = 0 \mid \beta, \delta_3 = +1) \cdot P(Y = 0 \mid \beta, \delta_4 = +2)$$

# Likelihood of the model given data

Assume we have observed answers to $I$ questions from $N$ persons, i.e., we have data $y$.

$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \cdots & y_{1,I} \\ y_{2,1} & y_{2,2} & \cdots & y_{2,I} \\ \cdots & & & \\ y_{N,1} & y_{N,2} & \cdots & y_{N,I} \end{array}$$

The task is to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood.

$$L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} P(y_{ni} \mid \beta_n, \delta_i) d\beta_n$$

... but this integral does not have a closed-form solution!

# Likelihood of the model given data

Assume we have observed answers to $I$ questions from $N$ persons, i.e., we have data $y$.

$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \dots & y_{1,I} \\ y_{2,1} & y_{2,2} & \dots & y_{2,I} \\ \dots & & & \\ y_{N,1} & y_{N,2} & \dots & y_{N,I} \end{array}$$

The task is to find model parameters $\sigma, \delta_1, \dots, \delta_I$ that maximize likelihood.

$$L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} P(y_{ni} \mid \beta_n, \delta_i) d\beta_n$$

... but this integral does not have a closed-form solution!

# Likelihood of the model given data

Assume we have observed answers to $I$ questions from $N$ persons, i.e., we have data $y$.

$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \dots & y_{1,I} \\ y_{2,1} & y_{2,2} & \dots & y_{2,I} \\ \dots & & & \\ y_{N,1} & y_{N,2} & \dots & y_{N,I} \end{array}$$

The task is to find model parameters $\sigma, \delta_1, \dots, \delta_I$ that maximize likelihood.

$$L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} P(y_{ni} \mid \beta_n, \delta_i) d\beta_n$$

... but this integral does not have a closed-form solution!

# Likelihood of the model given data

Assume we have observed answers to $I$ questions from $N$ persons, i.e., we have data $y$.

$$\begin{array}{cccc} y_{1,1} & y_{1,2} & \dots & y_{1,I} \\ y_{2,1} & y_{2,2} & \dots & y_{2,I} \\ \dots & & & \\ y_{N,1} & y_{N,2} & \dots & y_{N,I} \end{array}$$

The task is to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood.

$$L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} P(y_{ni} \mid \beta_n, \delta_i) d\beta_n$$

... but this integral does not have a closed-form solution!

# Likelihood of the model given data

Assume we have observed answers to $I$ questions from $N$ persons, i.e., we have data $y$.

$$
\begin{array}{cccc}
y_{1,1} & y_{1,2} & \cdots & y_{1,I} \\
y_{2,1} & y_{2,2} & \cdots & y_{2,I} \\
\cdots & & & \\
y_{N,1} & y_{N,2} & \cdots & y_{N,I}
\end{array}
$$

The task is to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood.

$$
L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} P(y_{ni} \mid \beta_n, \delta_i) d\beta_n
$$

... but this integral does not have a closed-form solution!

# Likelihood of the model given data

Assume we have observed answers to $I$ questions from $N$ persons, i.e., we have data $y$.

$$
\begin{array}{cccc}
y_{1,1} & y_{1,2} & \cdots & y_{1,I} \\
y_{2,1} & y_{2,2} & \cdots & y_{2,I} \\
\cdots & & & \\
y_{N,1} & y_{N,2} & \cdots & y_{N,I}
\end{array}
$$

The task is to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood.

$$
L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} P(y_{ni} \mid \beta_n, \delta_i) d\beta_n
$$

... but this integral does not have a closed-form solution!

# Likelihood of the model given data

Assume we have observed answers to $I$ questions from $N$ persons, i.e., we have data $y$.

$$
\begin{array}{cccc}
y_{1,1} & y_{1,2} & \cdots & y_{1,I} \\
y_{2,1} & y_{2,2} & \cdots & y_{2,I} \\
\cdots & & & \\
y_{N,1} & y_{N,2} & \cdots & y_{N,I}
\end{array}
$$

The task is to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood.

$$
L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} P(y_{ni} \mid \beta_n, \delta_i) d\beta_n
$$

... but this integral does not have a closed-form solution!

# Approximations to the integral

- Gaussian quadrature
- Laplace approximation
- Variational approximation

# Approximations to the integral

- Gaussian quadrature
- Laplace approximation
- Variational approximation

# Approximations to the integral

- Gaussian quadrature
- Laplace approximation
- Variational approximation

# Variational approximation

Let $h = \beta_n - \delta_i$.

$$P(Y_{n,i} = 1 \mid h) = \frac{\exp(h)}{1+\exp(h)}$$
$$= \frac{\exp(h/2)}{\exp(-h/2)+\exp(h/2)}$$
$$= \exp\left\{h/2 - \log\left[\exp(h/2)+\exp(-h/2)\right]\right\}$$
$$= \exp\left\{h/2 + f(h)\right\}$$

$f(h)$ is approximated by the first order Taylor expansion $\tilde{f}(h,\xi)$ in variable $h^2$ around point $\xi$.

$$P(Y_{n,i} = 1 \mid h) \approx \exp\left\{h/2 + \left[f(\xi) + \left.\frac{\partial f(h)}{\partial(h^2)}\right|_{h=\xi} \cdot (h^2 - \xi^2)\right]\right\}$$
$$= \exp\left\{h/2 + \tilde{f}(h,\xi)\right\}$$
$$= \tilde{P}(Y_{n,i} = 1 \mid h, \xi)$$

# Variational approximation

Let $h = \beta_n - \delta_i$.

$$P(Y_{n,i} = 1 \mid h) = \frac{\exp(h)}{1 + \exp(h)}$$
$$= \frac{\exp(h/2)}{\exp(-h/2) + \exp(h/2)}$$
$$= \exp\{h/2 - \log[\exp(h/2) + \exp(-h/2)]\}$$
$$= \exp\{h/2 + f(h)\}$$

$f(h)$ is approximated by the first order Taylor expansion $\tilde{f}(h, \xi)$ in variable $h^2$ around point $\xi$.

$$P(Y_{n,i} = 1 \mid h) \approx \exp\left\{h/2 + \left[f(\xi) + \frac{\partial f(h)}{\partial(h^2)}\bigg|_{h=\xi} \cdot (h^2 - \xi^2)\right]\right\}$$
$$= \exp\left\{h/2 + \tilde{f}(h, \xi)\right\}$$
$$= \tilde{P}(Y_{n,i} = 1 \mid h, \xi)$$

# Variational approximation

Let $h = \beta_n - \delta_i$.

$$
\begin{aligned}
P(Y_{n,i} = 1 \mid h) &= \frac{\exp(h)}{1 + \exp(h)} \\
&= \frac{\exp(h/2)}{\exp(-h/2) + \exp(h/2)} \\
&= \exp\left\{h/2 - \log\left[\exp(h/2) + \exp(-h/2)\right]\right\} \\
&= \exp\left\{h/2 + f(h)\right\}
\end{aligned}
$$

$f(h)$ is approximated by the first order Taylor expansion $\tilde{f}(h, \xi)$ in variable $h^2$ around point $\xi$.

$$
\begin{aligned}
P(Y_{n,i} = 1 \mid h) &\approx \exp\left\{h/2 + \left[f(\xi) + \left.\frac{\partial f(h)}{\partial(h^2)}\right|_{h=\xi} \cdot (h^2 - \xi^2)\right]\right\} \\
&= \exp\left\{h/2 + \tilde{f}(h, \xi)\right\} \\
&= \tilde{P}(Y_{n,i} = 1 \mid h, \xi)
\end{aligned}
$$

# Variational approximation

Let $h = \beta_n - \delta_i$.

$$
\begin{aligned}
P(Y_{n,i} = 1 \mid h) &= \frac{\exp(h)}{1 + \exp(h)} \\
&= \frac{\exp(h/2)}{\exp(-h/2) + \exp(h/2)} \\
&= \exp\left\{h/2 - \log\left[\exp(h/2) + \exp(-h/2)\right]\right\} \\
&= \exp\left\{h/2 + f(h)\right\}
\end{aligned}
$$

$f(h)$ is approximated by the first order Taylor expansion $\tilde{f}(h, \xi)$ in variable $h^2$ around point $\xi$.

$$
\begin{aligned}
P(Y_{n,i} = 1 \mid h) &\approx \exp\left\{h/2 + \left[f(\xi) + \left.\frac{\partial f(h)}{\partial (h^2)}\right|_{h=\xi} \cdot (h^2 - \xi^2)\right]\right\} \\
&= \exp\left\{h/2 + \tilde{f}(h, \xi)\right\} \\
&= \tilde{P}(Y_{n,i} = 1 \mid h, \xi)
\end{aligned}
$$

# Variational approximation

Let $h = \beta_n - \delta_i$.

$$
\begin{aligned}
P(Y_{n,i} = 1 \mid h) &= \frac{\exp(h)}{1+\exp(h)} \\
&= \frac{\exp(h/2)}{\exp(-h/2)+\exp(h/2)} \\
&= \exp\{h/2 - \log[\exp(h/2)+\exp(-h/2)]\} \\
&= \exp\{h/2 + f(h)\}
\end{aligned}
$$

$f(h)$ is approximated by the first order Taylor expansion $\tilde{f}(h,\xi)$ in variable $h^2$ around point $\xi$.

$$
\begin{aligned}
P(Y_{n,i} = 1 \mid h) &\approx \exp\left\{h/2 + \left[f(\xi) + \left.\frac{\partial f(h)}{\partial(h^2)}\right|_{h=\xi} \cdot (h^2 - \xi^2)\right]\right\} \\
&= \exp\left\{h/2 + \tilde{f}(h,\xi)\right\} \\
&= \tilde{P}(Y_{n,i} = 1 \mid h, \xi)
\end{aligned}
$$

# Variational approximation

Let $h = \beta_n - \delta_i$.

$$
\begin{aligned}
P(Y_{n,i} = 1 \mid h) &= \frac{\exp(h)}{1 + \exp(h)} \\
&= \frac{\exp(h/2)}{\exp(-h/2) + \exp(h/2)} \\
&= \exp\left\{h/2 - \log\left[\exp(h/2) + \exp(-h/2)\right]\right\} \\
&= \exp\left\{h/2 + f(h)\right\}
\end{aligned}
$$

$f(h)$ is approximated by the first order Taylor expansion $\tilde{f}(h,\xi)$ in variable $h^2$ around point $\xi$.

$$
\begin{aligned}
P(Y_{n,i} = 1 \mid h) &\approx \exp\left\{h/2 + \left[f(\xi) + \left.\frac{\partial f(h)}{\partial (h^2)}\right|_{h=\xi} \cdot (h^2 - \xi^2)\right]\right\} \\
&= \exp\left\{h/2 + \tilde{f}(h,\xi)\right\} \\
&= \tilde{P}(Y_{n,i} = 1 \mid h, \xi)
\end{aligned}
$$

# Variational approximation

Let $h = \beta_n - \delta_i$.

$$\begin{aligned} P(Y_{n,i} = 1 \mid h) &= \frac{\exp(h)}{1+\exp(h)} \\ &= \frac{\exp(h/2)}{\exp(-h/2)+\exp(h/2)} \\ &= \exp\left\{h/2 - \log\left[\exp(h/2)+\exp(-h/2)\right]\right\} \\ &= \exp\left\{h/2 + f(h)\right\} \end{aligned}$$

$f(h)$ is approximated by the first order Taylor expansion $\tilde{f}(h,\xi)$ in variable $h^2$ around point $\xi$.

$$\begin{aligned} P(Y_{n,i} = 1 \mid h) &\approx \exp\left\{h/2 + \left[f(\xi) + \left.\frac{\partial f(h)}{\partial (h^2)}\right|_{h=\xi} \cdot (h^2 - \xi^2)\right]\right\} \\ &= \exp\left\{h/2 + \tilde{f}(h,\xi)\right\} \\ &= \tilde{P}(Y_{n,i} = 1 \mid h, \xi) \end{aligned}$$

# Variational approximation

Let $h = \beta_n - \delta_i$.

$$\begin{aligned} P(Y_{n,i} = 1 \mid h) &= \frac{\exp(h)}{1 + \exp(h)} \\ &= \frac{\exp(h/2)}{\exp(-h/2) + \exp(h/2)} \\ &= \exp\left\{h/2 - \log\left[\exp(h/2) + \exp(-h/2)\right]\right\} \\ &= \exp\left\{h/2 + f(h)\right\} \end{aligned}$$

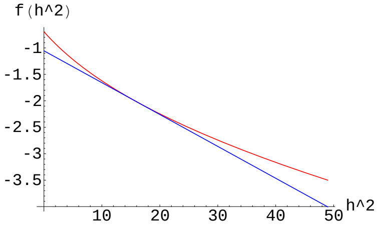
$f(h)$ is approximated by the first order Taylor expansion $\tilde{f}(h, \xi)$ in variable $h^2$ around point $\xi$.

$$\begin{aligned} P(Y_{n,i} = 1 \mid h) &\approx \exp\left\{h/2 + \left[f(\xi) + \left.\frac{\partial f(h)}{\partial (h^2)}\right|_{h=\xi} \cdot (h^2 - \xi^2)\right]\right\} \\ &= \exp\left\{h/2 + \tilde{f}(h, \xi)\right\} \\ &= \tilde{P}(Y_{n,i} = 1 \mid h, \xi) \end{aligned}$$

# Variational approximation

$$f(h) = \log\left[\exp(h/2) + \exp(-h/2)\right]$$

is a convex function in variable $x = h^2$

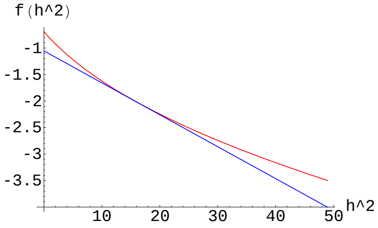



- Therefore, $f(h) \geq \tilde{f}(h, \xi)$ and, consequently, $\tilde{P}(Y_{n,i} = 1 \mid h, \xi)$ is a lower bound of $P(Y_{n,i} = 1 \mid h)$.
- $\tilde{f}(h, \xi)$ is a quadratic function of $\beta$ therefore $\tilde{P}(Y_{n,i} = 1 \mid h, \xi) = \exp\left\{h/2 + \tilde{f}(h, \xi)\right\}$ is proportional to a Gaussian distribution $\mathcal{N}(\nu, \tau^2)$

# Variational approximation

$$f(h) = \log\left[\exp(h/2) + \exp(-h/2)\right]$$

is a convex function in variable $x = h^2$

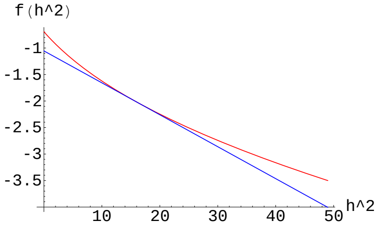



- Therefore, $f(h) \geq \tilde{f}(h, \xi)$ and, consequently, $\tilde{P}(Y_{n,i} = 1 \mid h, \xi)$ is a lower bound of $P(Y_{n,i} = 1 \mid h)$.
- $\tilde{f}(h, \xi)$ is a quadratic function of $\beta$ therefore $\tilde{P}(Y_{n,i} = 1 \mid h, \xi) = \exp\left\{h/2 + \tilde{f}(h, \xi)\right\}$ is proportional to a Gaussian distribution $\mathcal{N}(\nu, \tau^2)$

# Variational approximation

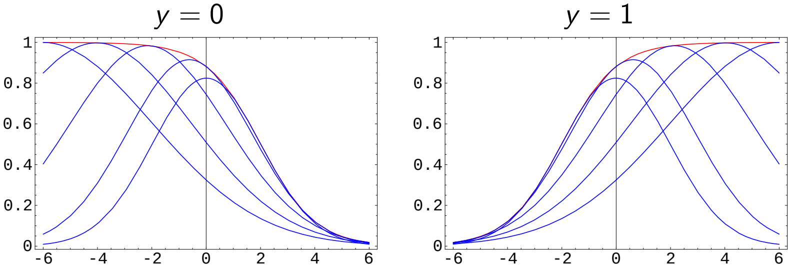
$$f(h) = \log\left[\exp(h/2) + \exp(-h/2)\right]$$

is a convex function in variable $x = h^2$




- Therefore, $f(h) \geq \tilde{f}(h,\xi)$ and, consequently, $\tilde{P}(Y_{n,i} = 1 \mid h, \xi)$ is a lower bound of $P(Y_{n,i} = 1 \mid h)$.
- $\tilde{f}(h,\xi)$ is a quadratic function of $\beta$ therefore $\tilde{P}(Y_{n,i} = 1 \mid h, \xi) = \exp\left\{h/2 + \tilde{f}(h,\xi)\right\}$ is proportional to a Gaussian distribution $\mathcal{N}(\nu, \tau^2)$

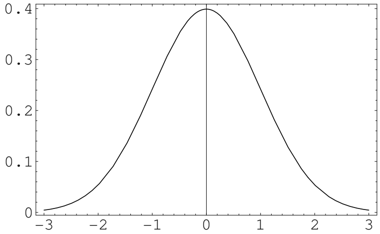
# Variational approximation of $P(Y_{n,i} = y \mid \delta, \beta)$ by $\tilde{P}(Y_{n,i} = y \mid \delta, \beta, \xi)$




$\xi = -6, -4, -2, 0, +2$ $\quad$ $\xi = -2, 0, +2, +4, +6$

# Computations with the Rasch model using a variational approximation with $\xi_i = 0.5$ for $i = 1, \ldots, 4$

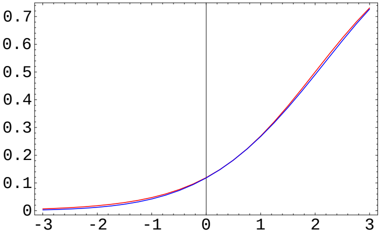
prior $\mathcal{N}_\beta(0,1)$



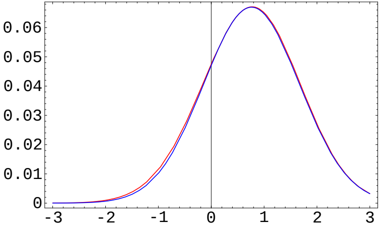

$$\mathcal{N}_\beta(0,1)$$

# Computations with the Rasch model using a variational approximation with $\xi_i = 0.5$ for $i = 1, \ldots, 4$

$P(Y = 1 \mid \beta, \delta_1 = -2)$

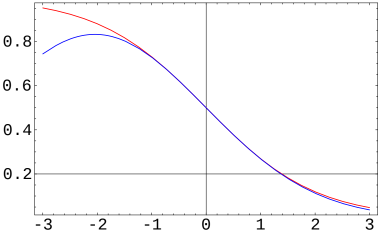


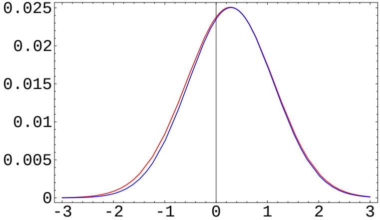

posterior




$$\mathcal{N}_\beta(0,1) \cdot \tilde{P}(Y = 1 \mid \beta, \delta_1 = -2)$$

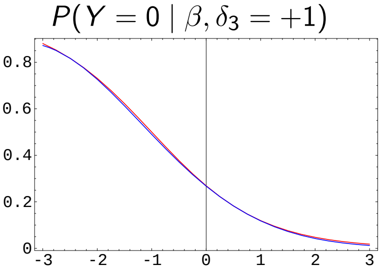
# Computations with the Rasch model using a variational approximation with $\xi_i = 0.5$ for $i = 1, \ldots, 4$

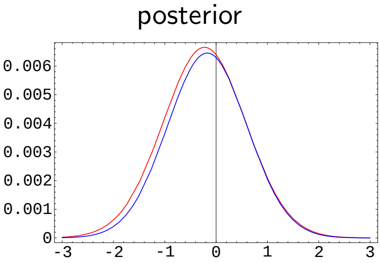



$$\mathcal{N}_\beta(0,1) \cdot \tilde{P}(Y = 1 \mid \beta, \delta_1 = -2) \cdot \tilde{P}(Y = 0 \mid \beta, \delta_2 = 0)$$

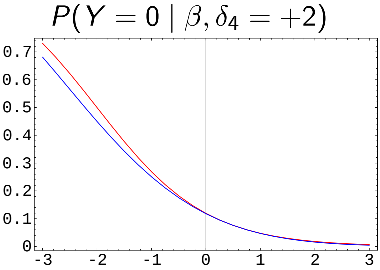
# Computations with the Rasch model using a variational approximation with $\xi_i = 0.5$ for $i = 1, \ldots, 4$

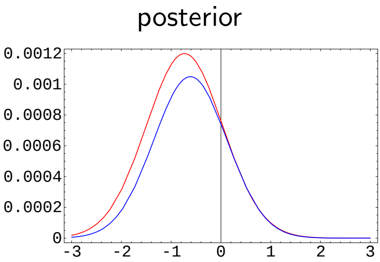



$$\mathcal{N}_\beta(0,1) \cdot \tilde{P}(Y = 1 \mid \beta, \delta_1 = -2) \cdot \tilde{P}(Y = 0 \mid \beta, \delta_2 = 0)$$
$$\cdot \tilde{P}(Y = 0 \mid \beta, \delta_3 = +1)$$

# Computations with the Rasch model using a variational approximation with $\xi_i = 0.5$ for $i = 1, \ldots, 4$




$$\mathcal{N}_\beta(0,1) \cdot \tilde{P}(Y = 1 \mid \beta, \delta_1 = -2) \cdot \tilde{P}(Y = 0 \mid \beta, \delta_2 = 0)$$
$$\cdot \tilde{P}(Y = 0 \mid \beta, \delta_3 = +1) \cdot \tilde{P}(Y = 0 \mid \beta, \delta_4 = +2)$$

# Optimization of the variational parameter $\xi$

- In the example we had the value of $\xi_i = 0.5$ fixed for $i = 1, 2, 3, 4$.
- But we can find the optimal value of $\xi_i$ for each $i$.
- The task was to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood. If we substitute the lower bounds $\tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i})$

$$L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} \tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i}) d\beta_n$$

... this integral has a closed-form solution!

- Since for any value of $\xi_i$ we have that $\tilde{P}(Y_i \mid \beta, \delta_i, \xi_i)$ is a lower bound, we can find the best value of $\xi_i$ as the value that maximizes approximated likelihood.
- For example, we can use the EM-algorithm to find optimal parameters.

# Optimization of the variational parameter $\xi$

- In the example we had the value of $\xi_i = 0.5$ fixed for $i = 1, 2, 3, 4$.
- But we can find the optimal value of $\xi_i$ for each $i$.
- The task was to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood. If we substitute the lower bounds $\tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i})$

$$L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} \tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i}) d\beta_n$$

... this integral has a closed-form solution!

- Since for any value of $\xi_i$ we have that $\tilde{P}(Y_i \mid \beta, \delta_i, \xi_i)$ is a lower bound, we can find the best value of $\xi_i$ as the value that maximizes approximated likelihood.
- For example, we can use the EM-algorithm to find optimal parameters.

# Optimization of the variational parameter $\xi$

- In the example we had the value of $\xi_i = 0.5$ fixed for $i = 1, 2, 3, 4$.
- But we can find the optimal value of $\xi_i$ for each $i$.
- The task was to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood. If we substitute the lower bounds $\tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i})$

$$L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} \tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i}) d\beta_n$$

... this integral has a closed-form solution!

- Since for any value of $\xi_i$ we have that $\tilde{P}(Y_i \mid \beta, \delta_i, \xi_i)$ is a lower bound, we can find the best value of $\xi_i$ as the value that maximizes approximated likelihood.
- For example, we can use the EM-algorithm to find optimal parameters.

# Optimization of the variational parameter $\xi$

- In the example we had the value of $\xi_i = 0.5$ fixed for $i = 1, 2, 3, 4$.
- But we can find the optimal value of $\xi_i$ for each $i$.
- The task was to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood. If we substitute the lower bounds $\tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i})$

$$L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} \tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i}) d\beta_n$$

... this integral has a closed-form solution!

- Since for any value of $\xi_i$ we have that $\tilde{P}(Y_i \mid \beta, \delta_i, \xi_i)$ is a lower bound, we can find the best value of $\xi_i$ as the value that maximizes approximated likelihood.
- For example, we can use the EM-algorithm to find optimal parameters.

# Optimization of the variational parameter $\xi$

- In the example we had the value of $\xi_i = 0.5$ fixed for $i = 1, 2, 3, 4$.
- But we can find the optimal value of $\xi_i$ for each $i$.
- The task was to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood. If we substitute the lower bounds $\tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i})$

$$L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} \tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i}) d\beta_n$$

... this integral has a closed-form solution!

- Since for any value of $\xi_i$ we have that $\tilde{P}(Y_i \mid \beta, \delta_i, \xi_i)$ is a lower bound, we can find the best value of $\xi_i$ as the value that maximizes approximated likelihood.
- For example, we can use the EM-algorithm to find optimal parameters.

# Optimization of the variational parameter $\xi$

- In the example we had the value of $\xi_i = 0.5$ fixed for $i = 1, 2, 3, 4$.
- But we can find the optimal value of $\xi_i$ for each $i$.
- The task was to find model parameters $\sigma, \delta_1, \ldots, \delta_I$ that maximize likelihood. If we substitute the lower bounds $\tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i})$

$$L = \prod_{n=1}^{N} \int \mathcal{N}_{\beta_n}(0, \sigma^2) \cdot \prod_{i=1}^{I} \tilde{P}(y_{n,i} \mid \beta_n, \delta_i, \xi_{n,i}) d\beta_n$$

... this integral has a closed-form solution!

- Since for any value of $\xi_i$ we have that $\tilde{P}(Y_i \mid \beta, \delta_i, \xi_i)$ is a lower bound, we can find the best value of $\xi_i$ as the value that maximizes approximated likelihood.
- For example, we can use the EM-algorithm to find optimal parameters.